



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Solomon, J. A., May, K. A. & Tyler, C. W. (2016). Inefficiency of orientation averaging: evidence for hybrid serial/parallel temporal integration. *Journal of Vision*, 16(1), 13. doi: 10.1167/16.1.13

This is the published version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/13014/>

**Link to published version:** <https://doi.org/10.1167/16.1.13>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---



# Inefficiency of orientation averaging: Evidence for hybrid serial/parallel temporal integration

Joshua A. Solomon

Centre for Applied Vision Research,  
City University London, UK



Keith A. May

Department of Computer Science,  
University College London, UK



Christopher W. Tyler

Centre for Applied Vision Research,  
City University London, UK



Intuition suggests that increased viewing time should allow for the accumulation of more visual information, but scant support for this idea has been found in studies of voluntary averaging, where observers are asked to make decisions based on perceived average size. In this paper we examine the dynamics of information accrual in an orientation-averaging task. With orientation (unlike intensive dimensions such as size), it is relatively safe to use an item's physical value as an approximation for its average perceived value. We displayed arrays containing eight iso-eccentric Gabor patterns, and asked six trained psychophysical observers to compare their average orientation with that of probe stimuli that were visible before, during, or only after the presentation of the Gabor array. From the relationship between orientation variance and human performance, we obtained estimates of *effective set size*, i.e., the number of items that an ideal observer would need to assess in order to estimate average orientation as well as our human observers did. We found that display duration had only a modest influence on effective set size. It rose from an average of  $\sim 2$  for 0.1-s displays to an average of  $\sim 3$  for 3.3-s displays. These results suggest that the visual computation is neither purely serial nor purely parallel. Computations of this nature can be made with a hybrid process that takes a series of subsamples of a few elements at a time.

abstracting certain statistical homogeneities" (p. 188). The goal of our research has been a better understanding of how visual statistics such as this are utilized in visual estimation, regardless of whether perceptual capacity is exceeded.

In this paper, we focus on "voluntary averaging" (Dakin, Bex, Cass, & Watt, 2009), a statistical summary of visual input that is distinct from crowding, in which the computation of textural statistics may be compulsory (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Specifically, we ask whether observers have access to textural or quasi-textural mechanisms that can process the feature content of multiple items in parallel, or whether observers must cognitively combine serial estimates from individual items in order to attain an estimate for the desired statistic (in our case, the average orientation in an array of Gabor patterns).

To quantify how well summary statistics like average orientation are calculated, we have adopted an Equivalent Noise (Nagaraja, 1964; Pelli, 1990; Dakin, 2001) framework for collecting and analyzing psychophysical data. Within this framework, there are two distinct limits on visual performance. The first is inefficiency, whereby observers do not utilize information relevant to their task. The second is internal noise, which decreases the fidelity with which stimuli are represented in the visual system. With an appropriate distribution, the addition of external noise (i.e., random perturbations of stimulus parameters) can mimic the effects of internal noise. For that reason, such external noise is known as "equivalent" noise when its variance matches that of the internal noise. Without any external noise, it is impossible to know whether performance is limited by inefficiency or internal noise. However, when the external noise is much greater than

## Introduction

Attneave (1954) noted, "When some portion of the visual field contains a quantity of information grossly in excess of the observer's perceptual capacity, he treats those components of information . . . as a statistician treats 'error variance,' averaging out particulars and

Citation: Solomon, J. A., May, K. A., & Tyler, C. W. (2016). Inefficiency of orientation averaging: Evidence for hybrid serial/parallel temporal integration. *Journal of Vision*, 16(1):13, 1–7, doi:10.1167/16.1.13.

doi: 10.1167/16.1.13

Received December 31, 2014; accepted November 24, 2015; published January 20, 2016

ISSN 1534-7362

This work is licensed under a Creative Commons Attribution 4.0 International License.



Downloaded From: <https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934836/> on 08/24/2018

| Experiment                                    | Paradigm                          | Display duration         | Set size             | Effective set size                  |
|---|-----------------------------------|--------------------------|----------------------|-------------------------------------|
| Dakin (2001)                                  | Classification w.r.t. vertical    | 0.10 s                   | $4 \leq N \leq 1024$ | $M \approx \sqrt{N}$                |
| Dakin et al. (2009) “Baseline”                | Classification w.r.t. vertical    | 0.15 s                   | $N = 6$              | $3 \leq M \leq 4$                   |
| Solomon (2010) Experiments 2 and 4            | 2-temporal-interval forced-choice | 0.15 s, with a 1.5-s ISI | $1 \leq N \leq 8$    | $M = \min\{N, x\}, 1 \leq x \leq 3$ |
| Allard & Cavanagh (2012)                      | Classification w.r.t. vertical    | 0.20 s                   | $N = 2, 4$           | $1.5 \leq M \leq 2.8$               |
| Tibber et al. (2015) Healthy control subjects | Classification w.r.t. vertical    | 0.40 s                   | $N = 100$            | $M = 2$ , on average                |

Table 1. Voluntary averaging of orientation.

the internal noise, the internal noise has a negligible effect, and efficiency can be computed from the ratio of human performance to that of an ideal observer.<sup>1</sup> Efficiency can be denoted  $M/N$ , where  $N$  represents the number of items on display and  $M$  represents the *effective set size*, i.e., the number of these items an otherwise-ideal observer would need to examine in order to perform as well as a human observer in high levels of external noise. The relationship between external noise, internal noise, effective set size, and performance is described by the noisy, inefficient observer model, a mathematical expression for which is provided at the beginning of the Results section.

Of course, the issue of parallel versus serial processing has been explored ad nauseam in the literature on visual search. After several decades and hundreds of papers, the field is now confident that some parallel processing is possible, but some tasks also require a serial deployment of attention from one group of items to another (e.g., Wolfe & Horowitz, 2004). If those parallel processes had fixed efficiencies, then it might be possible to infer the dynamics of attention from conventional measures, such as reaction time versus set size. However, there really is no reason to think that efficiency (or the effective set size) remains invariant with the number of items in a typical search display.<sup>2</sup>

Equivalent-noise analysis is a particularly useful tool with which to segregate the possibilities of parallel and serial mechanisms because it provides mathematical constraints on efficiency. Quite simply, if voluntary averaging were mediated by a purely serial process, i.e., one that estimates the orientations of individual items, one at a time, then (a) the effective set size should grow with the time available for processing the stimulus, and (b) it should be possible to prevent the serial process from having sufficient time to estimate the properties of more than one item in the array. The opposite of a purely serial process is a purely parallel process, which can estimate the orientations of multiple items, all at the same time. If voluntary averaging were mediated by such a process, then effective set size should remain constant with duration.

In this paper, we concentrate on orientation averaging, but the literature contains plenty of papers discussing the efficiencies of psychophysical decisions about such disparate visual features as luminance (e.g.,

Pelli, 1990), motion direction (Dakin, Mareschal, & Bex, 2005), and dot-matrix regularity (Morgan, Mareschal, Chubb, & Solomon, 2012). With some perceptual dimensions, there is a nonlinear mapping from the stimulus to the internal representation, and this needs to be modeled by a nonlinear transducer in the noisy, inefficient observer model; an incorrect form of transducer can lead to incorrect measures of efficiency. The advantage of using orientation in the present study is that, although we may not know exactly how apparent orientation is related to physical orientation, we can nonetheless expect that physically vertical things appear close to vertical and physically horizontal things appear close to horizontal. Thus, the psychophysical function for orientation may stray a little from the identity function (i.e., mapping every physical orientation to the identical apparent orientation), but it cannot stray too far.

With Table 1, we have attempted to compile a comprehensive list of the literature on voluntary averaging of orientation, from which estimates of effective set size are available. Dakin (2001) deserves credit for the popularity of this quantity, which he measured using both small and large arrays of Gabor patterns. His results contain a fair degree of scatter, but en masse, they suggested a relationship of the form  $M = N^p$ , where  $0.5 \leq p \leq 0.6$ . Later studies using small set sizes ( $N \leq 8$ ) seem to have produced results consistent with Dakin's ( $M \approx 2$ –3), but the one later study that used  $N = 100$  did not. Thus, Dakin's original study remains unique in its finding of large effective set sizes ( $M > 4$ ) for the voluntary averaging of orientation.

Marc Tibber (personal communication, December 4, 2014) suggested that practice might be necessary for high efficiency in voluntary averaging tasks. Support for this suggestion comes from comparing the methods of Dakin (2001), in which each observer performed more than 17,000 trials,<sup>3</sup> to those of Tibber et al. (2015), in which each observer completed a mere 150 orientation-averaging trials. We must stress, however, that practice does not seem to be a sufficient criterion for high efficiency, because Solomon (2010) describes one professional psychophysicist who competed 2,000 trials, yet achieved an effective set size no greater than 1.

It seemed plausible that performance in Solomon's (2010) task may have been hampered by the two-

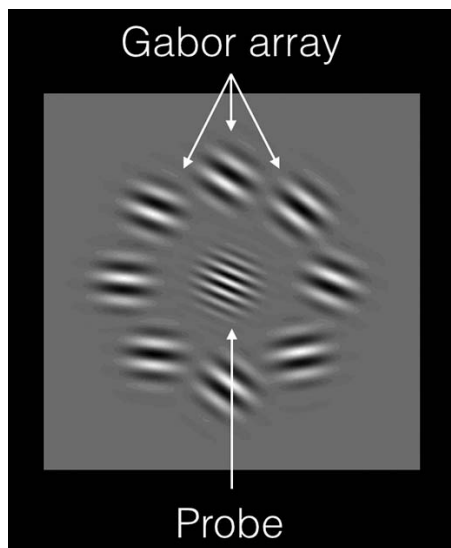


Figure 1. Stimulus layout. Observers were shown eight Gabor patterns of 3 cycles per degree in a ring of  $1.7^\circ$  radius around fixation, with orientations drawn from a Wrapped Normal distribution. The mean of the orientation distribution was random, and its standard deviation was either 0 or  $16^\circ$ . The central orientation probe was 6 cycles per degree.

interval forced-choice paradigm. In order to perform above chance levels, observers had to remember the average orientation of a briefly flashed Gabor array for 1.5 s, until the next Gabor array was flashed. In all of the other experiments summarized in Table 1, observers merely had to classify the average as clockwise or anticlockwise of vertical. It has already been established that the fidelity of memory for orientation decays faster than that for texture, Vernier alignment, contrast, spatial frequency, and the direction of motion (Pasternak & Greenlee, 2005). If Solomon's 1.5-s interstimulus interval (ISI) affected the precision with which individual orientations could be remembered, it also might have affected the efficiency of orientation averaging.

Consequently, we decided to manipulate memory load as well as display duration in our experiment. We had observers compare an array of Gabor patterns (see Figure 1) with the orientation of a different “probe” Gabor at fixation, presented at different times relative to the Gabor array in order to vary the memory load. Observers reported whether the probe was clockwise or anticlockwise of the mean across the circular array. The dependent variable was the just-classifiable angle (as clockwise or anticlockwise) between the array's expected orientation and the probe. The independent variables were the array's exposure duration, and whether the probe was displayed before, at the same time, or after the array.

Hypotheses regarding the effect of stimulus duration on efficiency were outlined previously. If voluntary

averaging were mediated by a purely serial process, then (a) efficiency should grow with the time available for processing the stimulus, and (b) it might be possible to prevent the serial process from having sufficient time to estimate the properties of more than one item in the array. The efficiency of a purely parallel process, on the other hand, should remain constant with duration. As for the other independent variable, we expected a general facilitation of performance when probes were exposed before the Gabor array (no memory load) and a general reduction in performance when probes were exposed after it (high memory load).

## Method

This experiment was approved by City University London's Senate Ethics panel, in conjunction with the EPSRC project “The Efficiency of Visual Statistics” (see Acknowledgments). All six observers (including authors JAS and KAM) had extensive experience with psychophysics. They were recruited from the Centre for Applied Vision Research, and provided written consent to participate in a noninvasive psychophysical experiment.

Stimuli were generated and responses were collected on a MacBook Pro computer. No attempt was made to correct for its native gamma function. The Psychtoolbox (Brainard, 1997; Pelli, 1997) was used for stimulus generation. Psychophysica (Watson & Solomon, 1997) was used for data analysis. Both codes are available upon request. Head positions were not restrained, but observers were asked to maintain a comfortable viewing distance ( $\sim 0.65$  m) for the duration of the experiment.

Stimulus arrays were composed of eight items, evenly distributed around an iso-eccentric circle (see Figure 1). At the viewing distance of 0.65 m, the radius of this circle subtended a visual angle of  $E = 1.7$  degrees (making the center-to-center separation of Gabors  $0.77E$ ) and there were 48 pixels per degree. Each item in the array was a Gabor pattern. It was the product of a sinusoidal luminance grating and a Gaussian blob. The grating had a spatial frequency of 3 cycles per degree and random spatial phase. The blob had a space constant (i.e., standard deviation  $\sigma$ ) of 0.25 degree of visual angle. Both grating and blob had maximum contrast. Spatial orientations were selected at random from a Wrapped Normal distribution with standard deviation,  $\sigma_G$ , either zero or  $16^\circ$ . The mean of this distribution (henceforth referred to as the “expected orientation”) was selected at random from a Uniform distribution over all orientations.

Equivalent noise analyses require a minimum of two measurements: performance in high levels of external



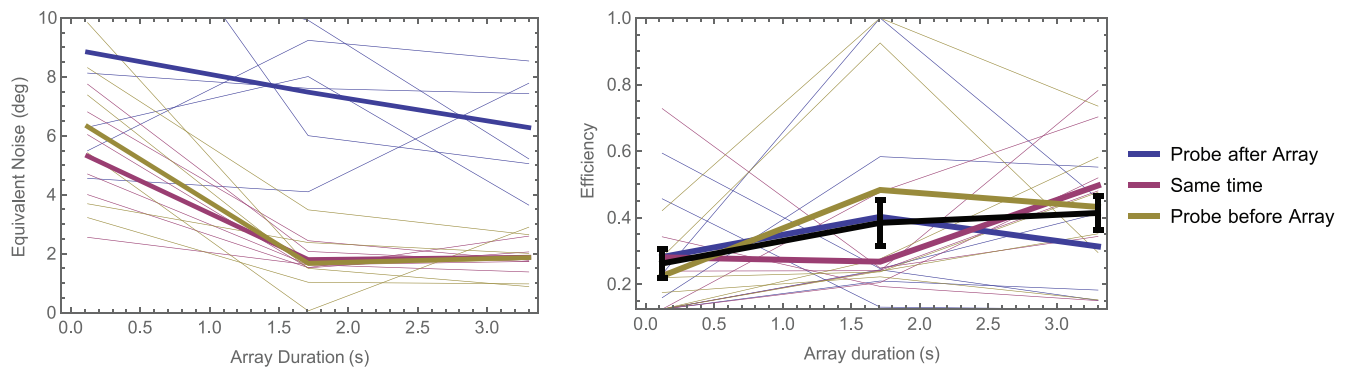


Figure 2. Panel a shows equivalent noise ( $\sigma_E$ ) and panel b shows efficiency ( $M/N$ ) versus display duration. Thin coloured lines illustrate results from individual observers. Thick colored lines illustrate mean values across observers. The black line in panel b illustrates the mean across observers and memory conditions. Error bars contain two standard errors of this mean.

noise and performance in low levels of (or zero) external noise. The value of  $16^\circ$  was selected because it promised to be larger than the standard deviation of internal noise, yet small enough to avoid the problem of orientation “wraparound” (Solomon, 2010).

The Gabor arrays remained visible for 0.1 s, 1.7 s, or 3.3 s. The probe Gabor appeared at the center of the circular array, where the observers were fixating. Its space constant was identical to that of the Gabors in the eight-item array, but its spatial frequency was twice as high. This choice was designed to discourage it from perceptually grouping with the array. Observers reported whether the probe was clockwise or anticlockwise of the array’s mean. The probe could appear 1.5 s before the array, it could appear at the same time as the array, or it could appear 1.5 s after the array had disappeared. It remained visible until the observer responded.

The angle ( $|\mu|$ ) between the probe and the expected orientation of the eight-item array was controlled by a QUEST staircase (Watson & Pelli, 1983) that was unique to each particular combination of display duration (0.1 s, 1.7 s, or 3.3 s), memory condition (probe before, at the same time as, or after the array), and level of external noise ( $\sigma_G = 0$  or  $\sigma_G = 16^\circ$ ).<sup>4</sup> The probe was clockwise or anticlockwise with equal probability (the observer’s task was to decide which), and never greater than  $37^\circ$  from the array’s expected mean. The staircases converged to 81%-correct thresholds. Different levels of external noise were interleaved within each block of trials, but display duration and memory condition were fixed, so as not to unduly handicap observers with uncertainty regarding stimulus dynamics. Each observer completed a minimum of either two blocks of 132 trials or three blocks of 88 trials in each of the nine conditions. (JAS and KAM completed a few more. QUEST was re-initialized at the beginning of each

block.) Consequently, Figure 2 summarizes more than 14,256 trials.

## Results

Our primary interest was in the efficiencies with which observers could discriminate clockwise from anticlockwise probes. To estimate those efficiencies we fit a simplified version of the noisy, inefficient observer model containing only early noise (Dakin, 2001; Solomon, 2010). Efficiency estimates from an alternative version, containing only late noise, would have been identical.<sup>5</sup> Specifically, we found the values of  $\sigma_E$  and  $M$  that maximized the joint likelihood of responses when the probability with which an observer responds “anticlockwise” is given by the formula:

$$\Pr(\text{“ACW”}) = \gamma + (1 - \gamma - \delta) \Phi \left[ \frac{\mu}{\sqrt{(\sigma_E^2 + \sigma_G^2)/M}} \right], \quad (1)$$

where  $\Phi$  denotes the standard normal cumulative distribution function. We assumed that lapse rates with anticlockwise and clockwise probes would be similar, and adopted equal values of  $\gamma$  and  $\delta$  for each observer, derived from the “easy” trials described in Footnote 4. Specifically, these values were 0.01 for observers JAS, JH, and TMP; and 0.04 for observers KAM, AJ, and CDC.

Remember that  $M$  represents the *effective set size*, i.e., the number of items an ideal observer would need to measure in order to estimate average orientations as well as our human observers. Consequently  $M \leq N$ , where  $N$  denotes the number of items in each set. The only further constraint placed on the model was that  $M \geq 1$ . When range effects, finger errors, invisible (or

unattended) stimuli, and perverse response strategies are eliminated, discrimination must be based on at least one item.

The threshold from each block of trials is illustrated in Supplementary Figure S1. To determine whether there was any effect of practice, thresholds were subjected to a two-way analysis of variance. There were 108 levels of the first factor, one for each unique combination of observer, display duration, memory condition, and external noise. The second factor was block number. Unsurprisingly, the main effect of the first factor was huge. It yielded an  $F$  ratio of  $F(1, 107) = 66.6$ ,  $p < 10^{-25}$ . The main effect of block number was nonsignificant,  $F(1, 3) = 0.4$ ,  $p > 0.75$ . In other words, we found no effect of practice in our experiment. Maximum-likelihood fits of the noisy, inefficient observer model (Equation 1) appear in Figure 2. Mean values across observers (thick lines) indicate that equivalent noise decreases and effective set size increases as the array duration increases.

First, we confirmed that there were significant individual differences between observers. This was achieved by comparing a pair of nested models using the generalized likelihood ratio test. The model with the fewest parameters had 55 free parameters: 54 of these set the early noise  $\sigma_E$  for each combination of six observers, three display durations, and three memory conditions, and there was a further free parameter that set the same effective set size  $M$  for each observer and condition. Against this model, we compared the fit of a 60-parameter model, which again had 54  $\sigma_E$  parameters but now had six  $M$  parameters, one for each observer. Because they are nested, the model with more parameters will always fit at least as well as the more-restricted model. To determine whether the fit is significantly better, we calculate the statistic,  $D$ , given by

$$D = 2\ln\left(\frac{L_2}{L_1}\right), \quad (2)$$

where  $L_2$  is the likelihood of the best-fitting model with more parameters, and  $L_1$  is the likelihood of the more restricted model. If the model with more parameters is no better (i.e., the null hypothesis is true), then  $D$  is distributed approximately as the  $\chi^2$  distribution with degrees of freedom given by the difference between the numbers of parameters in the two models (Mood, Graybill, & Boes, 1974, pp. 440–441). Therefore, a  $\chi^2$  test indicates whether the less-restricted model is significantly better. For the comparison between the two models described previously, we have  $\chi^2_{(5)} = 12.5$ ,  $p = 0.0001$ , indicating that the six observers were not all equally efficient. Graphically, this can be appreciated by the scatter of thin lines in Figure 2b.

Taking the significantly better, 60-parameter, model as our baseline, we then tested whether display

duration or memory condition had a significant effect by adding extra parameters to allow the efficiency to vary with these conditions. The first of these less-restricted models had 18  $M$  parameters, one for each combination of observer and display duration (giving 72 parameters in total). This model fit the data significantly better ( $\chi^2_{(12)} = 17.1$ ,  $p = 0.0006$ ), indicating that efficiency did in fact increase with display duration (black line in Figure 2b). The second of the less-restricted models had 18  $M$  parameters, one for each combination of observer and memory condition (again giving 72 parameters in total). This model did not fit significantly better than the 60-parameter baseline ( $\chi^2_{(12)} = 9.2$ ,  $p = 0.2$ ), indicating that memory condition did not significantly affect efficiency.

In the test of the effect of memory condition just described, we forced efficiency to be constant with respect to display duration for each observer. In a further test of the effect of memory condition, we allowed efficiency to vary with both observer and display duration. The more-restricted model had 72 parameters (54  $\sigma_E$  parameters and 18  $M$  parameters, one for each combination of observer and display condition), and the less-restricted model had 108 parameters (54  $\sigma_E$  parameters and 54  $M$  parameters, i.e., each combination of observer, display duration and memory condition had its own parameter for noise and efficiency). The less-restricted model did not fit significantly better ( $\chi^2_{(36)} = 7.8$ ,  $p = 0.999$ ). Fits of the 108-parameter model are illustrated in Supplementary Figure S1. Constraints on the effective set-sizes are illustrated in Supplementary Figure S2.

## Discussion

The results did not support the hypothesis of reduced efficiency with increased memory load. At first glance, this null result may seem hard to reconcile with experiments on visual working memory (VWM; e.g., Sims, Jacobs, & Knill, 2012), which utilize similar stimuli. However, capacity limits typically become manifest when VWM tasks require multiple display items to be encoded. Even in our “high memory load” condition, on the other hand, all observers needed to remember was a single statistic: the average orientation. Perhaps this is why we found that memory load did not affect our estimates of efficiency.

The results did support serial and, to a limited degree, parallel processes for orientation averaging. A serial process is supported because efficiency increased with stimulus duration. A parallel process is *consistent* with effective set sizes greater than 1 at the shortest duration, but there were no poststimulus masks in this

experiment. It is conceivable, therefore, that 0.125 s plus the duration of iconic memory (Sperling, 1960) provided enough time for a serial mechanism to utilize two items.

To appreciate how the visual system might compute average orientation in a manner that is neither purely serial nor purely parallel, consider the “Markovian subsampler” described by Gorea, Belkoura, and Solomon (2014). When information appears, at time  $t = 0$ , a subsample of maximum size  $m$  is selected and its average value, the baseline  $\hat{\mu}_1$ , is computed efficiently but noisily. We assume that the value of each item in the subsample is independently perturbed by the same stochastic process. This process manifests as early noise in fits of the noisy, inefficient observer model. Size  $m$  is said to be a maximum because there may be fewer items available in the display. Some  $\tau$  seconds later, another subsample is selected, and its mean  $\mu_2$  is computed with the same precision. A new baseline is then formed from the sum  $\hat{\mu}_2 = (1 - p_2)\hat{\mu}_1 + p_2\mu_2$ . The baseline continues to be updated every  $\tau$  seconds with newly selected subsamples until, at time  $t = T$ , the information disappears. For notational convenience we define  $S$  to be the total number of selected subsamples, i.e.,  $S = \lceil T/\tau \rceil$ . The final value of the baseline  $\hat{\mu}_S = (1 - p_S)\hat{\mu}_{S-1} + p_S\mu_S$  is then perturbed by another stochastic process, which manifests as late noise in fits of the NIO.

If  $m = 1$ , this Markovian subsampler will compute average orientation in a purely serial manner. Purely parallel computations require that  $m > 1$  and all subsamples must be identical. Regardless of the manner in which subsamples are selected, the baseline will be continually updated. Consequently a decrease in the total equivalent noise is consistent with serial, parallel, and hybrid versions of the Markovian subsampler.

Although the finding of an effective set size of 2 allows us to be confident that the orientations of more than one Gabor are considered in voluntary averaging, it does not tell us how those orientations are measured by the visual system. Parallel measurement of multiple orientations may involve the same computations thought to underlie local estimates of Gabor orientation (e.g., Graf, Kohn, Jazayeri, & Movshon, 2011). However, it is conceivable that those computations can be bypassed altogether by parallel-processing mechanisms with input from multiple items.

There is some evidence that local orientation estimates are combined under crowded conditions (Parkes et al., 2001), but previous results suggest that the efficiency of orientation averaging does not vary with Gabor separation (Solomon, 2010). Why not? One possibility is that the same mechanism is responsible for computing the average orientation of crowded and uncrowded Gabors. Averaging of crowded orientation

signals may be compulsory (Parkes et al., 2001), simply because input to this mechanism cannot be restricted to arbitrarily small regions of the visual field. Regardless whether orientation averaging is compulsory or voluntary, the current results strongly indicate that it is not very efficient.

*Keywords:* orientation, efficiency, averaging

## Acknowledgments

This research was sponsored by the Engineering and Physical Sciences Research Council (Grant EP/H033955/1). KAM is currently supported by a grant to Li Zhaoping from the Gatsby Charitable Foundation.

Commercial relationships: none.

Corresponding author: Joshua A. Solomon.

Email: J.A.Solomon@city.ac.uk

Address: Centre for Applied Vision Research, City University London, UK.

## Footnotes

<sup>1</sup> In the literature on luminance and contrast processing, this proportion has been called “calculation efficiency,” “central efficiency,” “sampling efficiency,” and “high-noise efficiency” (Pelli & Farrell, 1999). More recent literature (e.g., Pelli, Burns, Farell, & Moore-Page, 2006) favors the unmodified term “efficiency,” as does virtually all of the literature on orientation averaging, as summarized in Table 1.

<sup>2</sup> Palmer, Ames, and Lindsey (1993) advocated a wide separation between search items and spatial precues for manipulating the “relevant” set size for visual search. We agree that inferences regarding efficiency are indeed less tenuous when these methods are applied. At the same time, it seems safe to assume that uncued items are unprocessed only when the separation is sufficient to eliminate crowding. This precaution is rarely taken in studies of visual search.

<sup>3</sup> 64 trials/run  $\times$  (“at least”) 3 runs/data point  $\times$  7 data points/“subcondition”  $\times$  13 “subconditions.”

<sup>4</sup> To estimate lapse rates for each individual observer, the maximum angle ( $37^\circ$ ) between the probe and the expected orientation of the eight-item array was presented on each trial with probability 0.1, irrespective of QUEST.

<sup>5</sup> Any internal noise affecting estimation of the orientation of the probe would be indistinguishable from late noise. This formula is equivalent to one containing only late noise, because set size  $N$  is fixed.



## References

- Allard, R., & Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *Journal of Vision*, 12(11):6, 1–12, doi:10.1167/12.11.6. [PubMed] [Article]
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, 183–193.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A – Optics, Image Science, and Vision*, 18(5), 1016–1026.
- Dakin, S. C., Bex, P. J., Cass, J. R., & Watt, R. J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision*, 9(11):28, 1–16, doi:10.1167/9.11.28. [PubMed] [Article]
- Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*, 45, 3027–3049.
- Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, 14(9):22, 1–14, doi:10.1167/14.9.22. [PubMed] [Article]
- Graf, A. B. A., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neuroscience*, 14, 239–247.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. New York: McGraw-Hill.
- Morgan, M. J., Mareschal, I., Chubb, C., & Solomon, J. A. (2012). Perceived pattern regularity computed as a summary statistic: Implications for camouflage. *Proceedings of the Royal Society of London: Series B, Biological Sciences*, 279, 2754–2760.
- Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *Journal of the Optical Society of America*, 54, 950–955.
- Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1), 108–130.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6(2), 97–107.
- Pelli, D. G. (1990). The quantum efficiency of vision. In C. B. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3–24). Cambridge, UK: Cambridge University Press.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 427–442.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46, 4646–4674.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 16, 647–653.
- Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, 119, 807–830.
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, 10(14):19, 1–16, doi:10.1167/10.14.19. [PubMed] [Article]
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 72, (whole no. 11).
- Tibber, M. S., Anderson, E. J., Bobin, T., Carlin, P., Shergill, S. S., & Dakin, S. C. (2015). Local and global limits on visual processing in schizophrenia. *PLoS ONE*, 10(2), e0117951, doi:10.1371/journal.pone.0117951.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, 33, 113–120.
- Watson, A. B., & Solomon, J. A. (1997). Psychophysics: Mathematica notebooks for psychophysical experiments. *Spatial Vision*, 10, 447–466.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how they do it. *Nature Reviews Neuroscience*, 5, 1–7.